

The passage time distribution for a birth-and-death chain: Strong stationary duality gives a first stochastic proof

JAMES ALLEN FILL¹

Department of Applied Mathematics and Statistics

The Johns Hopkins University

jimfill@jhu.edu and <http://www.ams.jhu.edu/~fill/>

ABSTRACT

A well-known theorem usually attributed to Keilson states that, for an irreducible continuous-time birth-and-death chain on the nonnegative integers and any d , the passage time from state 0 to state d is distributed as a sum of d independent exponential random variables. Until now, no probabilistic proof of the theorem has been known. In this paper we use the theory of strong stationary duality to give a stochastic proof of a similar result for discrete-time birth-and-death chains and geometric random variables, and the continuous-time result (which can also be given a direct stochastic proof) then follows immediately. In both cases we link the parameters of the distributions to eigenvalue information about the chain. We also discuss how the continuous-time result leads to a proof of the Ray–Knight theorem.

Intimately related to the passage-time theorem is a theorem of Fill that any fastest strong stationary time T for an ergodic birth-and-death chain on $\{0, \dots, d\}$ in continuous time with generator G , started in state 0, is distributed as a sum of d independent exponential random variables whose rate parameters are the nonzero eigenvalues of $-G$. Our approach yields the first (sample-path) construction of such a T for which individual such exponentials summing to T can be explicitly identified.

AMS 2000 subject classifications. Primary 60J25; secondary 60J35, 60J10, 60G40.

Key words and phrases. Markov chains, birth-and-death chains, passage time, strong stationary duality, anti-dual, eigenvalues, stochastic monotonicity, Ray–Knight theorem.

Date. Revised May 18, 2009.

¹Research supported by NSF grant DMS-0406104, and by The Johns Hopkins University’s Acheson J. Duncan Fund for the Advancement of Research in Statistics.

1 Introduction and summary

A well-known theorem usually attributed to Keilson [12] (Theorem 5.1A, together with Remark 5.1B; see also Section 1 of [11]), but which—as pointed out by Laurent Saloff-Coste via Diaconis and Miclo [5]—can be traced back at least as far as Karlin and McGregor [10, equation (45)], states that, for an irreducible continuous-time birth-and-death chain on the nonnegative integers and any d , the passage time from state 0 to state d is distributed as a sum of d independent exponential random variables with distinct rate parameters. Keilson, like Karlin and McGregor, proves this result by analytical (non-probabilistic) means.

Modulo the distinctness of the rates, and with additional information (see, e.g., [2]) relating the exponential rates to spectral information about the chain, the theorem can be recast as follows.

Theorem 1.1. *Consider a continuous-time birth-and-death chain with generator G^* on the state space $\{0, \dots, d\}$ started at 0, suppose that d is an absorbing state, and suppose that the other birth rates λ_i^* , $0 \leq i \leq d-1$, and death rates μ_i^* , $1 \leq i \leq d-1$, are positive. Then the absorption time in state d is distributed as the sum of d independent exponential random variables whose rate parameters are the d nonzero eigenvalues of $-G^*$.*

There is an analogue for discrete time:

Theorem 1.2. *Consider a discrete-time birth-and-death chain with transition kernel P^* on the state space $\{0, \dots, d\}$ started at 0, suppose that d is an absorbing state, and suppose that the other birth probabilities p_i^* , $0 \leq i \leq d-1$, and death probabilities q_i^* , $1 \leq i \leq d-1$, are positive. Then the absorption time in state d has probability generating function*

$$u \mapsto \prod_{j=0}^{d-1} \left[\frac{(1 - \theta_j)u}{1 - \theta_j u} \right],$$

where $-1 \leq \theta_j < 1$ are the d non-unit eigenvalues of P^* .

In this paper we will give a stochastic proof of Theorem 1.2 under the additional hypothesis that all eigenvalues of P^* are (strictly) positive; as we shall see later (Lemma 2.4), this implies another condition key to our development, namely, that

$$p_{i-1}^* + q_i^* < 1, \quad 1 \leq i \leq d. \quad (1.1)$$

Whenever P^* has nonnegative eigenvalues, the conclusion of Theorem 1.2 simplifies:

The absorption time in state d is distributed as the sum of d independent geometric random variables whose failure probabilities are the non-unit eigenvalues of P^ .*

The special-case of Theorem 1.2 for positive eigenvalues establishes the theorem in general by the following argument (which is unusual, in that it is not often easy to relate characteristics of a chain to a “lazy” modification). Choose any $\varepsilon \in (0, 1/2)$ and apply the special case of Theorem 1.2 to the “lazy” kernel $P^*(\varepsilon) := (1 - \varepsilon)I + \varepsilon P^*$. Let T^*

and $T^*(\varepsilon)$ denote the respective absorption times for P^* and $P^*(\varepsilon)$. Then $T^*(\varepsilon)$ has probability generating function (pgf)

$$\mathbf{E} s^{T^*(\varepsilon)} = \prod_{j=0}^{d-1} \left[\frac{\varepsilon(1 - \theta_j)s}{1 - (1 - \varepsilon(1 - \theta_j))s} \right]. \quad (1.2)$$

But the conditional distribution of $T^*(\varepsilon)$ given T^* is negative binomial with parameters T^* and ε , so the pgf of $T^*(\varepsilon)$ can also be computed as

$$\mathbf{E} s^{T^*(\varepsilon)} = \mathbf{E} \mathbf{E} \left(s^{T^*(\varepsilon)} \middle| T^* \right) = \mathbf{E} \left(\frac{\varepsilon s}{1 - (1 - \varepsilon)s} \right)^{T^*}. \quad (1.3)$$

Equating (1.2) and (1.3) and letting $u := \varepsilon s / [1 - (1 - \varepsilon)s]$, we find, as desired,

$$\mathbf{E} u^{T^*} = \left[\frac{(1 - \theta_j)u}{1 - \theta_j u} \right].$$

Later in this section we explain how Theorem 1.1 follows from Theorem 1.2, but in Section 5 we will also outline a direct stochastic proof of Theorem 1.1.

Remark 1.3. (a) Theorems 1.1 and 1.2 are the starting point of an in-depth consideration of separation cut-off for birth-and-death chains in [6].

(b) By a simple perturbation argument, Theorems 1.1 and 1.2 extend to all birth-and-death chains for which the birth rates λ_i^* (respectively, birth probabilities p_i^*), $0 \leq i \leq d-1$, are positive.

(c) There is a stochastic interpretation of the pgf in Theorem 1.2 even when some of the eigenvalues are negative (see (4.23) in [4]), but we do not know a stochastic proof (i.e., a proof that proceeds by constructing random variables) in that case.

(d) The condition (1.1) is closely related to the notion of (stochastic) monotonicity. All continuous-time, but not all discrete-time, birth-and-death chains are monotone. In discrete time, monotonicity for a general chain is the requirement that the distributions $P^*(i, \cdot)$ in the successive rows of P^* be stochastically nondecreasing, i.e., that $\sum_{k>j} P^*(i, k)$ be nondecreasing in i for each j . As noted in [3], for a discrete-time birth-and-death chain P^* , monotonicity is equivalent to the condition

$$p_{i-1}^* + q_i^* \leq 1, \quad 1 \leq i \leq d.$$

We need only prove the discrete-time Theorem 1.2 (or even just the special case where P^* has positive eigenvalues), for then given a continuous-time birth-and-death generator G^* we can consider the discrete-time birth-and-death kernels $P^*(\varepsilon) := I + \varepsilon G^*$, where I denotes the identity matrix and $\varepsilon > 0$ is chosen sufficiently small that $P^*(\varepsilon)$ is nonnegative and has positive eigenvalues. Let $T(\varepsilon)$ and T denote the absorption times for $P^*(\varepsilon)$ and G^* , respectively. Then it is simple to check that $\varepsilon T(\varepsilon)$ converges in law to T ; indeed, for any $0 < t < \infty$ we have

$$\mathbf{P}(\varepsilon T(\varepsilon) \leq t) = (P^*(\varepsilon))^{\lfloor t/\varepsilon \rfloor}(0, d) \rightarrow (e^{tG^*})(0, d) = \mathbf{P}(T \leq t).$$

But the eigenvalues of $P^*(\varepsilon)$ and of $-G^*$ are simply related, and suitably scaled geometric random variables converge in law to exponentials, so Theorem 1.1 follows immediately.

The idea of our proof of Theorem 1.2 is simple: We show that the absorption time (call it T^*) of P^* has the same distribution as \hat{T} , where \hat{T} is the absorption time of a certain pure-birth chain \hat{P} whose holding probabilities are precisely the non-unit eigenvalues of P^* .

We do this by reviewing (in Section 2) and then employing the Diaconis and Fill [4] theory of strong stationary duality in discrete time. In brief, a given absorbing birth-and-death chain P^* satisfying (1.1) is the classical set-valued strong stationary dual (SSD) of some monotone birth-and-death chain P with the same eigenvalues; naturally enough, we will call P an “anti-dual” of P^* . But, if also the eigenvalues of P^* are nonnegative, then we show that this P (and indeed any ergodic birth-and-death chain with nonnegative eigenvalues) in turn also has a pure-birth SSD \hat{P} whose holding probabilities are precisely the non-unit eigenvalues of P . Since we argue that both duals are sharp (i.e., give rise to a stochastically minimal strong stationary time for the P -chain), the absorption time T^* of P^* has the same distribution as the absorption time \hat{T} of \hat{P} , and the latter distribution is manifestly the convolution of geometric distributions.

Remark 1.4. (a) Although our proof of Theorem 1.2 is stochastic, it leaves open [or, rather, left open—see part (c) of this remark] the question of whether the absorption time itself can be represented as an independent sum of explicit geometric random variables; the proof establishes only equality in distribution. The difficulty with our approach is that there can be many different stochastically minimal strong stationary times for a given chain.

(b) However, for either of the two steps of our argument we can give sample-path constructions relating the two chains (either P^* and P , or P and \hat{P}). This has already been carried out in detail for the first step in [4]. For the second step, what this means is that we can show how to watch the P -chain X run and contemporaneously construct from it a chain \hat{X} with kernel \hat{P} in such a way that the absorption time \hat{T} of \hat{P} is a fastest strong stationary time for X .

(c) Subsequent to the work leading to the present paper, Diaconis and Miclo [5] gave another stochastic proof of Theorem 1.1. Their proof, which provides an “intertwining” between the kernels P^* and \hat{P} (in our notation), yields a construction of exponentials summing to the absorption time, but the construction is, by their own estimation, “quite involved”. In a forthcoming paper [9], we will exhibit a much simpler such construction, with extensions to skip-free processes.

Section 2 is devoted to a brief review of strong stationary duality and a proof that any discrete-time birth-and-death kernel with positive eigenvalues satisfies (1.1). In Section 3, we construct P from P^* . In Section 4 we construct \hat{P} from P and (in Section 4.1) describe the sample-path construction discussed in Remark 1.4(b). In Section 5 for completeness we provide continuous-time analogs of our discrete-time auxiliary results, which we find interesting in their own right and which combine to give a direct stochastic proof of Theorem 1.1. Section 6 shows how to extend Theorem 1.1 from the hitting time of state d to the occupation-time vector for the states $\{0, \dots, d-1\}$

and connects the present paper with work of Kent [13] and the celebrated Ray–Knight theorem [17, 14].

2 A quick review of strong stationary duality

The main purpose of this background section is to review the theory of strong stationary duality only to the extent necessary to understand the proof of Theorem 1.2. For a more general and more detailed treatment, consult [4], especially Sections 2–4. To a reasonable extent, the notation of this paper matches that of [4]. Strong stationary duality has been used to bound mixing times of Markov chains and also to build perfect simulation algorithms [8].

2.1 Strong stationary duality in general

Let X be an ergodic (irreducible and aperiodic) Markov chain on a finite state space; call its stationary distribution π . A *strong stationary time* is a randomized stopping time T for X such that X_T has the distribution π and it independent of T . Aldous and Diaconis [1, Proposition 3.2] prove that for any such X there exists a fastest (i.e., stochastically minimal) strong stationary time, although it is well known that such a fastest time is not (generally) unique. (Such a fastest time is called a *time to stationarity* in [4], but this terminology has not been widely adopted and so will not be used here.)

A systematic approach to building strong stationary times is provided by the framework of strong stationary duality. The following specialization of the treatment in Section 2 of [4] (see especially Theorem 2.17 and Remark 2.39 there) is sufficient for our purposes.

Theorem 2.1. *Let π_0 and π_0^* be probability mass functions on $\{0, 1, \dots, d\}$, regarded as row vectors, and let P , P^* , and Λ be transition matrices on S . Assume that P is ergodic with stationary distribution π , that state d is absorbing for P^* , and that the row $\Lambda(d, \cdot)$ equals π . If (π_0^*, P^*) is a strong stationary dual of (π_0, P) with respect to the link Λ in the sense that*

$$\pi_0 = \pi_0^* \Lambda \quad \text{and} \quad \Lambda P = P^* \Lambda, \quad (2.1)$$

then there exists a bivariate Markov chain (X^, X) such that*

- (a) *X is marginally Markov with initial distribution π_0 and transition matrix P ;*
- (b) *X^* is marginally Markov with initial distribution π_0^* and transition matrix P^* ;*
- (c) *the absorption time T^* of X^* is a strong stationary time for X .*

Moreover, if $\Lambda(i, d) = 0$ for $i = 0, \dots, d-1$, then the dual is sharp in the sense that T^ is a fastest strong stationary time for X .*

Remark 2.2. In both our applications of Theorem 2.1 (Sections 3 and 4),

- (i) the initial distributions π_0 and π_0^* are both taken to be unit mass δ_0 at 0, and $\Lambda(0, \cdot) = \delta_0$, too, so only the second equation in (2.1) needs to be checked; and

- (ii) the link Λ is lower triangular, from which we observe that the corresponding dual is sharp and (if also the diagonal elements of Λ are all positive) that, given P , there is at most one stochastic matrix P^* satisfying (2.1), namely, $P^* = \Lambda P \Lambda^{-1}$.

2.2 Classical (set-valued) strong stationary duals

Let P be ergodic with stationary distribution π , and let H denote the corresponding cumulative distribution function (cdf):

$$H_j = \sum_{i \leq j} \pi_i.$$

Let Λ be the link of truncated stationary distributions:

$$\Lambda(x^*, x) = \mathbf{1}(x \leq x^*) \pi_x / H_{x^*}. \quad (2.2)$$

If P is a monotone birth-and-death chain (more generally, if P is arbitrary and the time reversal of P is monotone—see [4, Theorem 4.6]), then a dual P^* exists [and is sharp and unique by Remark 2.2(ii)]:

Theorem 2.3. *Let P be a monotone ergodic birth-and-death chain on $\{0, \dots, d\}$ with stationary cdf H . Then P has a sharp strong stationary dual P^* with respect to the link of truncated stationary distributions. The chain P^* is also birth-and-death, with death, hold, and birth probabilities (respectively)*

$$q_i^* = \frac{H_{i-1}}{H_i} p_i \quad r_i^* = 1 - (p_i + q_{i+1}), \quad p_i^* = \frac{H_{i+1}}{H_i} q_{i+1}. \quad (2.3)$$

See Sections 3–4 of [4] for an explanation as to why the dual in Theorem 2.3 is called “set-valued”; in this paper we shall refer to it as the “classical” SSD. The equations (2.3) reproduce [4, (4.18)].

2.3 Positivity of eigenvalues and stochastic monotonicity for birth-and-death chains

When we prove Theorem 1.2 assuming that P^* has positive eigenvalues, we will utilize the strengthened monotonicity condition (1.1). Part (a) of the following lemma provides justification.

Lemma 2.4. *Let P^* be the kernel of any birth-and-death chain on $\{0, \dots, d\}$.*

- (a) *If P^* has positive eigenvalues, then (1.1) holds.*
- (b) *If P^* has nonnegative eigenvalues, then P^* is monotone.*

Proof. (a) By perturbing P^* if necessary, we may assume that P^* is ergodic. Then P^* is diagonally similar to a positive definite matrix whose principal minor corresponding to rows and columns $i-1$ and i is $r_{i-1}^* r_i^* - p_{i-1}^* q_i^*$, so

$$0 < r_{i-1}^* r_i^* - p_{i-1}^* q_i^* \leq (1 - p_{i-1}^*)(1 - q_i^*) - p_{i-1}^* q_i^* = 1 - p_{i-1}^* - q_i^*.$$

- (b) This follows by perturbation from part (a). □

Remark 2.5. Both converse statements are false. For any given $d \geq 2$, the condition (1.1) does not imply nonnegativity of eigenvalues, not even for chains P^* satisfying the hypotheses of Theorem 1.2. An explicit counterexample for $d = 2$ is

$$P^* = \begin{bmatrix} 0.50 & 0.50 & 0 \\ 0.49 & 0.02 & 0.49 \\ 0 & 0 & 1 \end{bmatrix},$$

whose smallest eigenvalue is $(26 - \sqrt{3026})/100 < 0$. For general $d \geq 2$, perturb the direct sum of this counterexample with the identity matrix.

3 An anti-dual P of the given P^*

As discussed in Section 1, the main discrete-time theorem, Theorem 1.2, follows from the chief results, Theorems 3.1 and 4.2, of this section and the next.

Under the strengthened monotonicity condition (1.1) (with no assumption here about nonnegativity of the eigenvalues), the anti-dual construction of Theorem 3.1 exhibits the given chain (call its kernel P^*) as the classical SSD of another birth-and-death chain.

Theorem 3.1. *Consider a discrete-time birth-and-death chain P^* on $\{0, \dots, d\}$ started at 0, and suppose that d is an absorbing state. Write q_i^* , r_i^* , and p_i^* for its death, hold, and birth probabilities, respectively. Suppose that $p_i^* > 0$ for $0 \leq i \leq d-1$, that $q_i^* > 0$ for $1 \leq i \leq d-1$, and that $p_{i-1}^* + q_i^* < 1$ for $1 \leq i \leq d$. Then P^* is the classical (and hence sharp) SSD of some monotone ergodic birth-and-death kernel P on $\{0, \dots, d\}$.*

Proof. In light of Remark 2.2(i), we have dispensed with initial distributions. The claim is that P^* is related to some monotone ergodic P with stationary cdf H via (2.3). We will *begin* our proof by defining a suitable function H , and then we will construct P .

We inductively define a strictly increasing function $H : \{0, \dots, d\} \rightarrow (0, 1]$. Let $H_d := 1$, and define $H_{d-1} \in (0, 1)$ in (for now) arbitrary fashion. Having defined H_d, \dots, H_i (for some $1 \leq i \leq d-1$), choose the value of $H_{i-1} \in (0, H_i)$ so that

$$\left(\frac{H_i}{H_{i-1}} - 1 \right) q_i^* = \left(1 - \frac{H_i}{H_{i+1}} \right) p_i^*; \quad (3.1)$$

this is clearly possible since the right side of (3.1) is in $(0, 1)$ and the left side, as a function of the variable H_{i-1} , decreases from ∞ at $H_{i-1} = 0+$ to 0 at $H_{i-1} = H_i$. It is also clear that by choosing H_{d-1} sufficiently close to 1, we can make *all* the ratios H_i/H_{i-1} ($i = 1, \dots, d$) as (uniformly) close to 1 as we wish.

Next, define $q_0 := 0$,

$$p_0 := \left(1 - \frac{H_0}{H_1} \right) p_0^*, \quad (3.2)$$

and, for $1 \leq i \leq d$,

$$p_i := \frac{H_i}{H_{i-1}} q_i^*, \quad q_i := \frac{H_{i-1}}{H_i} p_{i-1}^*. \quad (3.3)$$

When the H -ratios are taken close enough to 1, then for $0 \leq i \leq d$ we have $p_i + q_i < 1$ and we define

$$r_i := 1 - p_i - q_i > 0.$$

The kernel P with death, hold, and birth probabilities q_i , r_i , and p_i is irreducible and aperiodic, and thus ergodic. To complete the proof, will also show

- (a) P is monotone (recall: equivalent to $p_i + q_{i+1} \leq 1$ for $0 \leq i \leq d-1$),
- (b) P has stationary cdf H , and
- (c) P^* is the classical SSD of P .

For (a) we simply observe, using (3.3) and (3.1), that

$$p_i + q_{i+1} = \frac{H_i}{H_{i-1}} q_i^* + \frac{H_i}{H_{i+1}} p_i^* = q_i^* + p_i^* \leq 1 \quad (3.4)$$

for $1 \leq i \leq d-1$; and similarly that

$$p_0 + q_1 = \left(1 - \frac{H_0}{H_1}\right) p_0^* + \frac{H_0}{H_1} p_0^* = p_0^* \leq 1.$$

For (b) we observe, again using (3.3) and (3.1), that the detailed balance condition

$$(H_i - H_{i-1})p_i = (H_i - H_{i-1})\frac{H_i}{H_{i-1}}q_i^* = (H_{i+1} - H_i)\frac{H_i}{H_{i+1}}p_i^* = (H_{i+1} - H_i)q_{i+1}$$

holds for $1 \leq i \leq d-1$; by (3.2) and (3.3), it also holds for $i = 0$:

$$H_0 p_0 = (H_1 - H_0)\frac{H_0}{H_1}p_0^* = (H_1 - H_0)q_1.$$

For (c), we simply verify that (2.3) holds: for $0 \leq i \leq d$ (with $H_{-1} := 0$), from (3.3) and (3.4),

$$\frac{H_{i-1}}{H_i} p_i = q_i^*, \quad \frac{H_{i+1}}{H_i} q_{i+1} = p_i^*, \quad p_i + q_{i+1} = q_i^* + p_i^* = 1 - r_i^*. \quad \square \quad (3.5)$$

Remark 3.2. Once the value of H_{d-1} is chosen, the definitions of H and P are forced; indeed, if the detailed balance condition and (3.5) are to hold, then we must have (3.1)–(3.3).

4 A pure birth “spectral” dual of P

In this section we construct a sharp pure birth dual \hat{P} for any ergodic birth-and-death chain P on $\{0, \dots, d\}$ with nonnegative eigenvalues started in state 0. When this construction is applied in the proof of Theorem 1.2 to the chain P resulting from P^* by application of Theorem 3.1, assuming nonnegativity of the eigenvalues of P^* yields the required nonnegativity of the eigenvalues of P in Theorem 4.2; indeed, as noted in Remark 2.2(ii), the matrices P and P^* are similar.

Our construction of the pure birth dual specializes a SSD construction of Matthews [15] for general reversible chains with nonnegative eigenvalues; that construction is closely related to the spectral decomposition of the transition matrix. For completeness and the reader's convenience, and because for birth-and-death chains (a) we can give a more streamlined presentation with minimal reference to eigenvectors and (b) we wish to establish the new result that the resulting dual is sharp, we do not presume familiarity with [15].

To set up our construction we need some notation. Let P be an ergodic birth-and-death chain on $\{0, \dots, d\}$ with stationary probability mass function π (note that π is everywhere positive) and nonnegative eigenvalues, say $0 \leq \theta_0 \leq \theta_1 \leq \dots \leq \theta_{d-1} < \theta_d = 1$. (It is well known [12] [4, Theorem 4.20] that the eigenvalues are all distinct, but we will not need this fact.) Let I denote the identity matrix and define

$$Q_k := (1 - \theta_0)^{-1} \cdots (1 - \theta_{k-1})^{-1} (P - \theta_0 I) \cdots (P - \theta_{k-1} I), \quad k = 0, \dots, d, \quad (4.1)$$

with the natural convention $Q_0 := I$. Note that for $k = 0, \dots, d-1$ we have

$$Q_k P = \theta_k Q_k + (1 - \theta_k) Q_{k+1}. \quad (4.2)$$

Lemma 4.1. *The matrices Q_k are all stochastic, and every row of Q_d equals π .*

Proof. For the first assertion it is clear that the rows of Q_k all sum to 1, so the only question is whether Q_k is nonnegative. But $P = D^{-1/2} S D^{1/2}$, where $D = \text{diag}(\pi)$ and S is symmetric, so the nonnegativity of Q_k follows from that of

$$S_k := (S - \theta_0 I) \cdots (S - \theta_{k-1} I),$$

which in turn is an immediate consequence of (the rather nontrivial) Theorem 3.2 in [16] using only that S is nonnegative and symmetric.

For the second assertion, write

$$S = \sum_{r=0}^d \theta_r u_r u_r^T,$$

where the column vectors u_0, \dots, u_d form an orthogonal matrix and u_d has i th entry $\sqrt{\pi_i}$. Then, as noted at (2.6) of [16],

$$S_k = \sum_{r=k}^d \left[\prod_{t=0}^{k-1} (\theta_r - \theta_t) \right] u_r u_r^T.$$

In particular, $S_d = (1 - \theta_0) \cdots (1 - \theta_{d-1}) u_d u_d^T$, so every row of Q_d equals π . \square

Now let δ_0 denote unit mass at 0 (regarded as a row vector), and define the probability mass functions

$$\lambda_k := \delta_0 Q_k, \quad k = 0, \dots, d. \quad (4.3)$$

Let $\widehat{\Lambda}$ [so named to distinguish it from the classic link Λ of (2.2)] be the lower-triangular square matrix with successive rows $\lambda_0, \dots, \lambda_d$, and define \widehat{P} to be the pure-birth chain

transition matrix on $\{0, \dots, d\}$ with holding probability θ_i at state i for $i = 0, \dots, d$; that is,

$$\hat{p}_{ij} := \begin{cases} \theta_i & \text{if } j = i \\ 1 - \theta_i & \text{if } j = i + 1 \\ 0 & \text{otherwise.} \end{cases} \quad (4.4)$$

Theorem 4.2. *Let P be an ergodic birth-and-death chain on $\{0, \dots, d\}$ with nonnegative eigenvalues. In the above notation, \hat{P} is a sharp strong stationary dual of P with respect to the link $\hat{\Lambda}$.*

Proof. We have again dispensed with initial distributions by Remark 2.2(i). The desired equation $\hat{\Lambda}P = \hat{P}\hat{\Lambda}$ is equivalent to

$$\lambda_k P = \theta_k \lambda_k + (1 - \theta_k) \lambda_{k+1}, \quad k = 0, \dots, d-1; \quad \lambda_d P = \lambda_d,$$

which is true because $\lambda_d = \pi$ and, for $k = 0, \dots, d-1$,

$$\lambda_k P = \delta_0 Q_k P = \theta_k \lambda_k + (1 - \theta_k) \lambda_{k+1}$$

by (4.2). The SSD is sharp because $\hat{\Lambda}$ is lower triangular; recall Remark 2.2(ii). \square

Remark 4.3. Lemma 4.1 is interesting and, as we have now seen, gives rise to the construction of a new “spectral” SSD for a certain subclass of monotone birth-and-death chains, namely, chains with nonnegative eigenvalues [recall Lemma 2.4(b)]. But for the proof of Theorem 1.2 one could make do without the nonnegativity of the matrix $\hat{\Lambda}$ by taking the approach of Matthews [15] and considering the chain P started in a suitable mixture of δ_0 and the stationary distribution π . We omit further details.

4.1 Sample-path construction of the spectral dual

Let X be an ergodic birth-and-death chain on $\{0, \dots, d\}$ with kernel P having nonnegative eigenvalues, assume $X_0 = 0$, and let T be any fastest strong stationary time for X . Independent of interest in Theorems 1.1 and 1.2, Theorem 4.2 gives the first stochastic interpretation of the individual geometrics in the representation of the distribution of T as a convolution of geometric distributions. In this subsection we carry this result one step further by showing how to construct, sample path by sample path, a particular fastest strong stationary time \hat{T} which is the sum of explicitly identified independent geometric random variables.

The idea is simple. Theorem 4.2 shows that \hat{P} of (4.4) is an “algebraic” dual of P in the sense that the matrix-equation $\hat{\Lambda}P = \hat{P}\hat{\Lambda}$ holds. But whenever algebraic duality holds for any finite-state ergodic chain with respect to any link ($\hat{\Lambda}$ in our case), Section 2.4 of [4] shows explicitly how to construct, from X and independent randomness, a dual Markov chain (\hat{X} in our case, with kernel \hat{P}) such that the absorption time \hat{T} of \hat{X} is a strong stationary time for X ; since $\hat{\Lambda}$ is lower triangular, \hat{T} will be stochastically optimal. So to describe our construction of \hat{X} (and hence \hat{T}) we need only specialize the construction of [4, Section 2.4] [see especially (2.36) there].

The chain X starts with $X_0 = 0$ and we set $\hat{X}_0 = 0$. Inductively, we will have $\hat{\Lambda}(\hat{X}_t, X_t) > 0$ (and so $X_t \leq \hat{X}_t$) at all times t . The value we construct for \hat{X}_t depends only on the values of \hat{X}_{t-1} and X_t and independent randomness. Indeed, given $\hat{X}_{t-1} = \hat{x}$ and $X_t = y$, if $y \leq \hat{x}$ then our construction sets $\hat{X}_t = \hat{x} + 1$ with probability

$$\frac{\hat{P}(\hat{x}, \hat{x} + 1) \hat{\Lambda}(\hat{x} + 1, y)}{(\hat{P} \hat{\Lambda})(\hat{x}, y)} = \frac{(1 - \theta_{\hat{x}}) \hat{\Lambda}(\hat{x} + 1, y)}{\theta_{\hat{x}} \hat{\Lambda}(\hat{x}, y) + (1 - \theta_{\hat{x}}) \hat{\Lambda}(\hat{x} + 1, y)} = \frac{(1 - \theta_{\hat{x}}) Q_{\hat{x}+1}(0, y)}{(Q_{\hat{x}} P)(0, y)} \quad (4.5)$$

and $\hat{X}_t = \hat{x}$ with the complementary probability; if $y = \hat{x} + 1$ (which is the only other possibility, since $y = X_t \leq X_{t-1} + 1 \leq \hat{x} + 1$ by induction), then we set $\hat{X}_t = \hat{x} + 1$ with certainty.

The independent geometric random variables, with sum \hat{T} , are the waiting times between successive births in the chain \hat{X} we have built. Thus it is no longer true that the individual geometric distributions “have no known interpretation in terms of the underlying [ergodic] birth and death chain” [6, Section 4, Remark 1]; likewise, for continuous time consult Section 5.1 herein.

Example 4.4. Consider the well-studied Ehrenfest chain, with holding probability $1/2$:

$$q_i = \frac{i}{2d}, \quad r_i = \frac{1}{2}, \quad p_i = \frac{d-i}{2d}, \quad i = 0, \dots, d.$$

The eigenvalues are $\theta_i \equiv i/d$. A straightforward proof by induction using (4.3) and (4.2) confirms that λ_k is the binomial distribution with parameters k and $1/2$:

$$\hat{\Lambda}(\hat{x}, x) \equiv \binom{\hat{x}}{x} 2^{-\hat{x}}. \quad (4.6)$$

Thus the probability (4.5) reduces to

$$\frac{(d - \hat{x})(\hat{x} + 1)}{2\hat{x}(\hat{x} + 1 - y) + (d - \hat{x})(\hat{x} + 1)}.$$

The chain we have described lifts naturally to random walk on the set \mathbf{Z}_2^d of binary d -tuples whereby one of the d coordinates is chosen uniformly at random and its entry is then replaced randomly by 0 or 1. It is interesting to note that the sharp pure-birth SSD chain constructed in this example does *not* correspond to the well-known “coordinate-checking” sharp SSD (see Example 3.2 of [4]). Indeed, expressed in the birth-and-death chain domain, the coordinate-checking dual is a pure-birth chain, call it \hat{X}' , such that the construction of \hat{X}'_t depends not only on \hat{X}'_{t-1} and X_t but also on X_{t-1} . The construction rules are that if $\hat{X}'_{t-1} = \hat{x}$, $X_{t-1} = x$, and $X_t = y$, then \hat{X}'_t is set to $\hat{x} + 1$ with probability

$$0 \text{ if } y = x - 1, \quad 1 - \frac{\hat{x}}{d} \text{ if } y = x, \quad \text{and} \quad \frac{d - \hat{x}}{d - x} \text{ if } y = x + 1,$$

and otherwise \hat{X}'_t holds at \hat{x} . Both duals correspond to the same link (4.6) and the (marginal) transition kernels for \hat{X} and \hat{X}' are the same, but the bivariate constructions of (\hat{X}, X) and (\hat{X}', X) are different.

The freedom for such differences was noted in [4, Remark 2.23(c)] and exploited in the creation of an interruptible perfect simulation algorithm (see [8, Remark 9.8]). In fact, \widehat{X}' (when lifted to \mathbf{Z}_2^d) corresponds to the construction used in [8]. An advantage of the \widehat{X} -construction of the present paper is that it allows (both in our Ehrenfest example and in general) for holding probabilities that are arbitrary (subject to nonnegativity of eigenvalues); in the paragraph containing (4.5), all that changes when a weighted average of the transition kernel and the identity matrix is taken are the eigenvalues $\theta_0, \dots, \theta_{d-1}$.

5 Continuous-time analogs of other results

As discussed in Section 1, the continuous-time Theorem 1.1 follows immediately from the discrete-time Theorem 1.2. Another way to prove Theorem 1.1 is to repeat the proof of Theorem 1.2 by establishing continuous-time analogs (namely, the next three results) of the auxiliary results (Theorem 3.1, Lemma 4.1, and Theorem 4.2) in the preceding two sections; we find these interesting in their own right. The continuous-time results are easy to prove utilizing the continuous-time SSD theory of [7], either by repeating the discrete-time proofs or by applying the discrete-time results to the appropriate kernel $P^*(\varepsilon) = I + \varepsilon G^*$ or $P(\varepsilon) = I + \varepsilon G$, with $\varepsilon > 0$ chosen sufficiently small to meet the hypotheses of those results; so we state the results without proof.

In Section 5.1 we will present the analog of Section 4.1 for continuous time.

Here, first, is the analog of Theorem 3.1.

Theorem 5.1. *Consider a continuous-time birth-and-death chain with generator G^* on $\{0, \dots, d\}$ started at 0, and suppose that d is an absorbing state. Write μ_i^* and λ_i^* for its death and birth rates, respectively. Suppose that $\lambda_i^* > 0$ for $0 \leq i \leq d-1$ and that $\mu_i^* > 0$ for $1 \leq i \leq d-1$. Then G^* is the classical set-valued (and hence sharp) SSD of some ergodic birth-and-death generator G on $\{0, \dots, d\}$.*

To set up the second result we need a little notation. Let G be the generator of a continuous-time ergodic birth-and-death chain on $\{0, \dots, d\}$ with stationary probability mass function π and eigenvalues $\nu_0 \geq \nu_1 \geq \dots \geq \nu_{d-1} > \nu_d = 0$ for $-G$. (Again, we don't need the fact [12] that the eigenvalues are distinct.) Define

$$Q_k := \nu_0^{-1} \cdots \nu_{k-1}^{-1} (G + \nu_0 I) \cdots (G + \nu_{k-1} I), \quad k = 0, \dots, d, \quad (5.1)$$

with the natural convention $Q_0 := I$.

Lemma 5.2. *The matrices Q_k are all stochastic, and every row of Q_d equals π .*

Now define $\widehat{\Lambda}$ in terms of the Q_k 's as in the paragraph preceding Theorem 4.2, and let \widehat{G} be the pure-birth generator on $\{0, \dots, d\}$ with birth rate ν_i at state i for $i = 0, \dots, d$.

Theorem 5.3. *Let G be the generator of an ergodic birth-and-death chain on $\{0, \dots, d\}$. In the above notation, \widehat{G} is a sharp strong stationary dual of G with respect to the link $\widehat{\Lambda}$:*

$$\widehat{\Lambda}G = \widehat{G}\widehat{\Lambda}.$$

5.1 Sample-path construction of the continuous-time spectral dual

Let X be an ergodic continuous-time birth-and-death chain on $\{0, \dots, d\}$, adopt all the notation of Section 5 thus far, and assume $X_0 = 0$. In this subsection by a routine application of Section 2.3 of [7] we give a simple sample-path construction of a “spectral dual” pure birth chain \hat{X} with generator \hat{G} as described just before Theorem 5.3; its absorption time \hat{T} is then a fastest strong stationary time for X and the independent exponential random variables with sum \hat{T} are simply the waiting times for the successive births for \hat{X} . We thus obtain a stochastic proof, with explicit identification of individual exponential random variables, of Theorem 5 in [7].

The chain X starts with $X(0) = 0$ and we set $\hat{X}(0) = 0$. Let $n \geq 1$ and suppose that \hat{X} has been constructed up through the epoch τ_{n-1} of the $(n-1)$ st transition for the bivariate process (\hat{X}, X) ; here $\tau_0 := 0$. We describe next, in terms of an exponential random variable \hat{V} , how to define τ_n and $\hat{X}(\tau_n)$; we will have $\hat{\Lambda}(\hat{X}(\tau_n), X(\tau_n)) > 0$ and hence $X(\tau_n) \leq \hat{X}(\tau_n)$. Write (\hat{x}, x) for the value of (\hat{X}, X) at time τ_{n-1} ; by induction we have $\hat{\Lambda}(\hat{x}, x) > 0$.

Let \hat{V}_n be exponentially distributed with rate

$$r = \nu_{\hat{x}} \hat{\Lambda}(\hat{x} + 1, x) / \hat{\Lambda}(\hat{x}, x), \quad (5.2)$$

independent of $\hat{V}_1, \dots, \hat{V}_{n-1}$ and the chain X . Consider two (independent) exponential waiting times begun at epoch τ_{n-1} : a first for the next transition of the chain X , and a second with rate r . How we proceed breaks into two cases:

- (i) If the first waiting time is smaller than the second, then τ_n is the epoch of this next transition for X and we set $\hat{X}(\tau_n) = \hat{x} = \hat{X}(\tau_{n-1})$ (with certainty) except in one circumstance: if $X(\tau_n) = \hat{x} + 1$, then we set $\hat{X}(\tau_n) = \hat{x} + 1$, too.
- (ii) If the second waiting time is smaller, then $\tau_n = \tau_{n-1} + \hat{V}_n$ and we set $\hat{X}(\tau_n) = \hat{x} + 1$.

Example 5.4. Consider the continuous-time version of the Ehrenfest chain with death rates $\mu_i \equiv i$ and birth rates $\lambda_i \equiv d - i$, $0 \leq i \leq d$; the eigenvalues are $\nu_i \equiv 2(d - i)$. Then $\hat{\Lambda}$ is again the link (4.6) of binomial distributions, and the rate (5.2) reduces to

$$r = \frac{(d - \hat{x})(\hat{x} + 1)}{\hat{x} + 1 - x}.$$

6 Occupation times and connection with Ray–Knight Theorem

Our final section utilizes work of Kent [13]; see the historical note at the end of Section 1 of [5] for closely related material. We show how to extend the continuous-time Theorem 1.1 from the hitting time of state d first to the occupation-time vector for the states $\{0, \dots, d-1\}$ and then to the local time of Brownian motion, thereby proving the Ray–Knight theorem [17, 14].

6.1 From hitting time to occupation times

Consider a continuous-time irreducible birth-and-death chain with generator G^* . It is then immediate from the Karlin–McGregor theorem (Theorem 1.1) that the hitting time T^* of state d has Laplace transform

$$\mathbf{E} e^{-uT^*} = \frac{\det(-G_0)}{\det(-G_0 + uI)}, \quad (6.1)$$

with G_0 obtained from G^* by leaving off the last row and column.

Equation (6.1) gives the distribution of the total time elapsed before the chain hits state d . But how is that time apportioned to the states $0, \dots, d-1$? This question can be answered from (6.1) using a neat trick of Kent [13] [see the last sentence of his Remark (1) on page 164]. To find the multivariate distribution of the occupation-time vector $\mathbf{T} = (T_0, T_1, \dots, T_{d-1})$, where T_i denotes the occupation time of (i.e., amount of time spent in) state i , it of course suffices to compute the value $\mathbf{E} e^{-\langle \mathbf{u}, \mathbf{T} \rangle}$ of the Laplace transform for any vector $\mathbf{u} = (u_0, \dots, u_{d-1})$ with strictly positive entries. But the distribution of the random variable $\langle \mathbf{u}, \mathbf{T} \rangle = \sum u_i T_i$ is that of the time to absorption for the time-changed generator $G_{\mathbf{u}}$ (say) obtained by dividing the i th row of G^* by u_i for $i = 0, \dots, d-1$. Therefore, by (6.1) and the scaling property of determinants,

$$\mathbf{E} e^{-\langle \mathbf{u}, \mathbf{T} \rangle} = \frac{\det(-G_{\mathbf{u}})}{\det(-G_{\mathbf{u}} + I)} = \frac{\det(-G_0)}{\det(-G_0 + U)},$$

where $U := \text{diag}(u_0, \dots, u_{d-1})$.

6.2 From occupation times to the Ray–Knight Theorem

Call the stationary distribution π . Then the matrix $S := D(-G_0)D^{-1}$ is (strictly) positive definite, where $D := \text{diag}(\sqrt{\pi})$. Let $\Sigma := \frac{1}{2}S^{-1}$. By direct calculation, \mathbf{T} has the same law as $\mathbf{Y} + \mathbf{Z}$, where \mathbf{Y} and \mathbf{Z} are independent random vectors with the same law and \mathbf{Y} is the coordinate-wise square of a Gaussian random vector $\mathbf{V} \sim N(0, \Sigma)$.

Kent [13] uses and extends this “double derivation” of $\mathcal{L}(\mathbf{T})$ to prove the theorem of Ray [17] and Knight [14] expressing the local time of Brownian motion as the sum of two independent 2-dimensional Bessel processes (i.e., as the sum of two independent squared Brownian motions).

Acknowledgments. We thank Persi Diaconis for helpful discussions, and Raymond Nung-Sing Sze and Chi-Kwong Li for pointing out the reference [16].

References

- [1] Aldous, D. and Diaconis, P. Strong uniform times and finite random walks. *Adv. in Appl. Math.* **8** (1987), 69–97.
- [2] Brown, M. and Shao, Y. S. Identifying coefficients in the spectral representation for first passage time distributions. *Probab. Eng. Inform. Sci.* **1** (1987), 69–74.

- [3] Cox, J. T. and Rösler, U. A duality relation for entrance and exit laws for Markov monotone Markov processes. *Ann. Probab.* **13** (1985), 558–565.
- [4] Diaconis, P. and Fill, J. Strong stationary times via a new form of duality. *Ann. Probab.* **18** (1990), 1483–1522.
- [5] Diaconis, P. and Miclo, L. On times to quasi-stationarity for birth and death processes. *J. Theoret. Probab.* (2009), to appear.
- [6] Diaconis, P. and Saloff-Coste, L. Separation cut-offs for birth and death chains. *Ann. Appl. Probab.* **16** (2006), 2098–2122.
- [7] Fill, J. A. Strong stationary duality for continuous-time Markov chains. Part I: Theory. *J. Theoret. Probab.* **5** (1992), 45–70.
- [8] Fill, J. A. An interruptible algorithm for perfect sampling via Markov chains. *Ann. Appl. Probab.* **8** (1998), 131–162.
- [9] Fill, J. A. On hitting times and fastest strong stationary times for skip-free and more general chains. *J. Theoret. Probab.* (2009), to appear.
- [10] Karlin, S. and McGregor, J. Coincidence properties of birth and death processes. *Pacific J. Math.* **9** (1959), 1109–1140.
- [11] Keilson, J. Log-concavity and log-convexity in passage time densities for of diffusion and birth-death processes. *J. Appl. Probab.* **8** (1971), 391–398.
- [12] Keilson, J. *Markov Chain Models—Rarity and Exponentiality*. Springer, New York, 1979.
- [13] Kent, J. T. The appearance of a multivariate exponential distribution in sojourn times for birth-death and diffusion processes. In *Probability, statistics and analysis*, volume 79 of *London Math. Soc. Lecture Note Ser.*, pages 161–179, Cambridge Univ. Press, Cambridge, 1983.
- [14] Knight, F. B. Random walks and a sojourn density process of Brownian motion. *Trans. Amer. Math. Soc.* **109** (1963), 56–86.
- [15] Matthews, P. Strong stationary times and eigenvalues. *J. Appl. Probab.* **29** (1992), 228–233.
- [16] Micchelli, C. A. and Willoughby, R. A. On functions which preserve the class of Stieltjes matrices. *Lin. Alg. Appl.* **23** (1979), 141–156.
- [17] Ray, D. Sojourn times of diffusion processes. *Illinois J. Math.* **7** (1963), 615–630.